

APPLICATION OF KALMAN FILTERING TO TRACK AND VERTEX FITTING

R. FRÜHWIRTH

Institut für Hochenergiephysik der Österreichischen Akademie der Wissenschaften, Vienna, Austria

Received 30 June 1987

Recently iterative procedures have been proposed for track and vertex fitting in counter experiments. We show that the proper theoretical framework for these procedures is the theory of linear filtering, in particular the Kalman filter. Using results from filtering theory we confirm and extend the previous results. We also discuss the detection of outliers and of secondary vertices.

1. Introduction

The “progressive” method of track fitting [1,2] which is used in the data analysis program of the DELPHI collaboration has some substantial advantages as compared with other methods, in particular the “global” method [2]:

- It is suitable for combined track finding and track fitting.
- No large matrices have to be inverted and the number of computations increases linearly with the number of measurements in the track. Therefore it is fast even in the presence of multiple scattering and many measurements.
- The estimated track parameters follow closely the physical track.
- The linear approximation of the track model needs to be valid only over a short range.

The progressive method has, however, one fundamental drawback: The track parameters are known with optimal precision only after the last step of the fit, i.e. usually at the inner end point of the track or track segment. In the presence of multiple scattering this has several consequences:

- Predictions into detectors further outwards are not optimal.
- The power of discrimination between measurements which may all belong to the track is rather poor at the begin of the track fit.
- Since the aim of the track fit is the optimal knowledge of the track parameters as close to the vertex as possible, the fit has to proceed towards the interaction region. This effectively prohibits the use of the progressive method in the forward region of the DELPHI detector [3], where it is natural to proceed from the TPC with its good track element towards the forward drift chambers [3].

It is the purpose of this note to show that these

difficulties can be overcome by applying techniques of linear filtering to track fitting. This has been proposed already some years ago [4], without success, as it seems, as the method was never used in practice.

In the framework of filtering theory the progressive method can be regarded as an extended Kalman filter. The smoothing part of the Kalman filter will be seen to be a very useful complement which solves the problems mentioned above and makes the progressive method a powerful, flexible and efficient tool not only for track fitting, but also for the computation of optimal predictions and interpolations, for outlier detection and rejection, and for merging of track segments.

In section 2 we clarify the relations between track fitting and filtering. In section 3 we present the main properties of the linear Kalman filter. In section 4 we discuss the application of the extended Kalman filter and the corresponding smoother to track fitting and track element merging. For the interested reader, the detection of outliers is investigated in a formal way in section 5. In section 6 we show that also the recursive vertex fit proposed in ref. [2] can be understood as an iterated Kalman filter. Finally, in section 7, we discuss the detection of secondary vertices.

2. Relations between track fitting and filtering

Track fitting deals with the estimation of track parameters; filtering deals with the analysis of (linear) dynamic systems [5,6]. We can apply filtering techniques to track fitting if we regard a track in space as a dynamic system. This can be done quite naturally by identifying the state vector of the dynamic system with a vector x of 5 parameters uniquely describing the track in each point of its trajectory. This state vector x can be written as a function of a suitable coordinate, e.g. z :

$$x = x(z).$$

The evolution of the state vector \mathbf{x} as a function of z can be described by a set of differential equations. For practical purposes, however, it is sufficient to consider the state vector only in a discrete set of points, e.g. in the intersection points $\mathbf{x}(z_k)$ of the track with the detectors. Thus we obtain a discrete dynamic system, the evolution of which can be described by a simple system equation:

$$\mathbf{x}(z_k) \equiv \mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1}. \quad (1)$$

\mathbf{f}_{k-1} is the track propagator from detector $k-1$ to detector k ; the random variable \mathbf{w}_{k-1} incorporates a random disturbance of the track between z_{k-1} and z_k due to multiple scattering. \mathbf{w}_{k-1} is called the process noise in the terminology of dynamic systems.

The state vector is normally not observed directly. Generally speaking, \mathbf{m}_k , the quantities measured by detector k , are functions of the state vector, corrupted by a measurement noise ϵ_k . This is described by the measurement equation:

$$\mathbf{m}_k = \mathbf{h}_k(\mathbf{x}_k) + \epsilon_k \quad (2)$$

We assume in the following that all \mathbf{w}_k and all ϵ_k are independent, unbiased and of finite variance.

In the simplest case, both the functions \mathbf{f}_k and the functions \mathbf{h}_k are linear. We then speak of a discrete linear dynamic system. In the next section we discuss the analysis of such a system, in particular the estimation of the state vector (track parameters).

3. Main properties of the discrete linear Kalman filter

The theory of the Kalman filter is described in many textbooks (e.g. refs. [5–8]). In its linear form the Kalman filter is the optimal recursive estimator of the state vector of a (discrete) linear dynamic system. In such a system the evolution of the state vector is described by a linear transformation plus a random disturbance \mathbf{w} , which is the process noise:

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}. \quad (3)$$

The measurements are linear functions of the state vector:

$$\mathbf{m}_k = \mathbf{H}_k\mathbf{x}_k + \epsilon_k. \quad (4)$$

By assumption all \mathbf{w}_k and all ϵ_k are independent and have a mean value of zero.

The evolution of the state vector as indicated by the index k may proceed in space, as in the case of track fitting in a detector, or in time, as in radar tracking of a spacecraft (in fact the origin of the method), or along a dimensionless integer, as in the common vertex fit of several tracks.

There are three types of operations to be performed

in the analysis of a dynamic system (here described in terms of “time”):

- *Filtering* is the estimation of the “present” state vector, based upon all “past” measurements.
- *Prediction* is the estimation of the state vector at a “future” time.
- *Smoothing* is the estimation of the state vector at some time in the “past” based on all measurements taken up to the “present” time.

The Kalman filter is the optimum solution of these three problems in the sense that it minimizes the mean square estimation error. If \mathbf{w}_k and ϵ_k are Gaussian random variables, the Kalman filter is *the* optimal filter; no nonlinear filter can do better. In other cases it is simply the optimal linear filter.

We give now the formulae for prediction, filtering and smoothing, with the following notations and assumptions:

System equation:

$$\mathbf{x}_{k,i} = \mathbf{F}_{k-1}\mathbf{x}_{k-1,i} + \mathbf{w}_{k-1}. \quad (5)$$

Measurement equation:

$$\mathbf{m}_k = \mathbf{H}_k\mathbf{x}_{k,i} + \epsilon_k, \quad (6)$$

$$E\{\mathbf{w}_k\} = 0, \text{ cov}\{\mathbf{w}_k\} = \mathbf{Q}_k,$$

$$E\{\epsilon_k\} = 0, \text{ cov}\{\epsilon_k\} = \mathbf{V}_k = \mathbf{G}_k^{-1},$$

with

$\mathbf{x}_{k,i}$ = true value of the state vector at time k ;

\mathbf{x}_k^i = estimate of $\mathbf{x}_{k,i}$, using measurements up to time i ($i < k$: prediction, $i = k$: filtered estimate, $i > k$: smoothed estimate), \mathbf{x}_k^k is simply written as \mathbf{x}_k ;

$$\mathbf{C}_k^i = \text{cov}\{\mathbf{x}_k^i - \mathbf{x}_{k,i}\};$$

$$\mathbf{r}_k^i = \text{residual } \mathbf{m}_k - \mathbf{H}_k\mathbf{x}_k^i;$$

$$\mathbf{R}_k^i = \text{cov}\{\mathbf{r}_k^i\}.$$

Prediction:

Extrapolation of the state vector:

$$\mathbf{x}_k^{k-1} = \mathbf{F}_{k-1}\mathbf{x}_{k-1}.$$

Extrapolation of the covariance matrix:

$$\mathbf{C}_k^{k-1} = \mathbf{F}_{k-1}\mathbf{C}_{k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}.$$

Residuals of predictions:

$$\mathbf{r}_k^{k-1} = \mathbf{m}_k - \mathbf{H}_k\mathbf{x}_k^{k-1}. \quad (7)$$

Covariance matrix of predicted residuals:

$$\mathbf{R}_k^{k-1} = \mathbf{V}_k + \mathbf{H}_k\mathbf{C}_k^{k-1}\mathbf{H}_k^T.$$

Filtering (gain matrix formalism):

Update of the state vector:

$$\mathbf{x}_k = \mathbf{x}_k^{k-1} + \mathbf{K}_k(\mathbf{m}_k - \mathbf{H}_k\mathbf{x}_k^{k-1}).$$

Kalman gain matrix:

$$\begin{aligned} \mathbf{K}_k &= \mathbf{C}_k^{k-1}\mathbf{H}_k^T(\mathbf{V}_k + \mathbf{H}_k\mathbf{C}_k^{k-1}\mathbf{H}_k^T)^{-1} \\ &= \mathbf{C}_k\mathbf{H}_k^T\mathbf{G}_k. \end{aligned}$$

Update of the covariance matrix:

$$\mathbf{C}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{C}_k^{k-1}. \quad (8a)$$

Filtered residuals:

$$\mathbf{r}_k = \mathbf{m}_k - \mathbf{H}_k \mathbf{x}_k = (\mathbf{I} - \mathbf{H}_k \mathbf{K}_k) \mathbf{r}_k^{k-1}.$$

Covariance matrix of filtered residuals:

$$\mathbf{R}_k = (\mathbf{I} - \mathbf{H}_k \mathbf{K}_k) \mathbf{V}_k = \mathbf{V}_k - \mathbf{H}_k \mathbf{C}_k \mathbf{H}_k^T.$$

χ^2 increment:

$$\chi_+^2 = \mathbf{r}_k^T \mathbf{R}_k^{-1} \mathbf{r}_k.$$

χ^2 update:

$$\chi_k^2 = \chi_{k-1}^2 + \chi_+^2.$$

Filtering (weighted means formalism):

Update of the state vector:

$$\mathbf{x}_k = \mathbf{C}_k \left[(\mathbf{C}_k^{k-1})^{-1} \mathbf{x}_k^{k-1} + \mathbf{H}_k^T \mathbf{G}_k \mathbf{m}_k \right].$$

Update of the covariance matrix:

$$\mathbf{C}_k = \left[(\mathbf{C}_k^{k-1})^{-1} + \mathbf{H}_k^T \mathbf{G}_k \mathbf{H}_k \right]^{-1}. \quad (8b)$$

χ^2 increment:

$$\chi_+^2 = \mathbf{r}_k^T \mathbf{G}_k \mathbf{r}_k + (\mathbf{x}_k - \mathbf{x}_k^{k-1})^T (\mathbf{C}_k^{k-1})^{-1} (\mathbf{x}_k - \mathbf{x}_k^{k-1}).$$

χ^2 update:

$$\chi_k^2 = \chi_{k-1}^2 + \chi_+^2.$$

Smoothing:

Smoothed state vector:

$$\mathbf{x}_k^n = \mathbf{x}_k + \mathbf{A}_k (\mathbf{x}_{k+1}^n - \mathbf{x}_{k+1}^k).$$

Smother gain matrix:

$$\mathbf{A}_k = \mathbf{C}_k \mathbf{F}_k^T (\mathbf{C}_{k+1}^k)^{-1}.$$

Covariance matrix of the smoothed state vector:

$$\mathbf{C}_k^n = \mathbf{C}_k + \mathbf{A}_k (\mathbf{C}_{k+1}^n - \mathbf{C}_{k+1}^k) \mathbf{A}_k^T. \quad (9)$$

Smoothed residuals:

$$\mathbf{r}_k^n = \mathbf{r}_k - \mathbf{H}_k (\mathbf{x}_k^n - \mathbf{x}_k) = \mathbf{m}_k - \mathbf{H}_k \mathbf{x}_k^n.$$

Covariance matrix of smoothed residuals:

$$\mathbf{R}_k^n = \mathbf{R}_k - \mathbf{H}_k \mathbf{A}_k (\mathbf{C}_{k+1}^n - \mathbf{C}_{k+1}^k) \mathbf{A}_k^T \mathbf{H}_k^T = \mathbf{V}_k - \mathbf{H}_k \mathbf{C}_k^n \mathbf{H}_k^T.$$

We make the following observations:

(1) The gain matrix formalism and the weighted means formalism of the filter are equivalent. The choice between the two depends on the dimensions of the state vector and the measurement vector. If the dimension of the state vector is small, the computation by weighted means is usually more efficient.

(2) The filtered estimate of the state vector is unbiased and has minimum variance among all linear estimates using the same set of measurements. For Gaussian process noise and measurement errors it is efficient. The same is true for the smoothed estimates. Therefore

the Kalman filter with a subsequent smoothing is equivalent to a global linear least-squares fit which takes into account all correlations arising from the process noise.

(3) The computation time of the filter is proportional to the number of detectors and depends (in the weighted means formalism) very little on the number of measurements per detector. If the intermediate results of the filter are retained the smoother consists only of a few matrix multiplications and is thus very fast.

(4) \mathbf{C}_k may also be expressed by the formula:

$$\mathbf{C}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{C}_k^{k-1} (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)^T + \mathbf{K}_k \mathbf{V}_k \mathbf{K}_k^T.$$

One can show that this form is computationally superior to the form given above in eq. (8a), although it consumes more computer time (ref. [5], p. 305).

(5) If there is no process noise ($\mathbf{Q}_k = 0$), smoothing is equivalent to back extrapolation, as can be verified directly from the smoother equations.

(6) Inspection of the covariance matrix update equations gives the following results, which are intuitively obvious: The variance of the filtered state vector is smaller than the variance of the predicted state vector (information from the measurement \mathbf{m}_k); the mean squared filtered residual is smaller than the mean squared predicted residual (the state vector is pulled towards the measurement); the variance of the smoothed state vector is smaller than the variance of the filtered state vector (information from all measurements); the mean squared smoothed residual is larger than the mean squared filtered residual (the state vector is pulled towards the true value).

(7) The filtered residual vectors (also called innovations) are uncorrelated, in the Gaussian case even independent. This is a characteristic property of the Kalman filter. It also proves the χ^2 update formula.

4. Application to track fitting and track element merging

In the presence of a magnetic field the track propagator (the system eq. (1)) is nonlinear. On the other hand, the measurement eq. (2) can usually be made linear by an appropriate choice of the state vector. In order to apply the concepts of linear filtering to track fitting, the track propagator has to be approximated by a linear function. This is done in the usual way by replacing \mathbf{f}_k by the first two terms of its Taylor expansion:

$$\mathbf{f}_k(\mathbf{x}_k^*) = \mathbf{f}_k(\mathbf{x}_k) + \mathbf{F}_k(\mathbf{x}_k^* - \mathbf{x}_k), \quad (10a)$$

$$\mathbf{F}_k = \partial \mathbf{f}_k / \partial \mathbf{x}_k. \quad (10b)$$

As the point of expansion we choose of course the filtered estimate \mathbf{x}_k . Apart from the state vector extrapolation, which now reads

$$\mathbf{x}_k^{k-1} = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}), \quad (11)$$

the prediction, filter and smoother equations remain the same. This procedure is called an extended Kalman filter.

The covariance matrix of the process noise, \mathbf{Q}_k , is computed in the same way as in the global fit, by integrating the effects of multiple scattering between detectors k and $k + 1$ (see, e.g., ref. [9]). Consequently, the amount of computing spent in the evaluation of derivatives and covariance matrices is exactly the same for both approaches. The only difference is in the inversion of covariance matrices.

Note that in the absence of multiple scattering smoothing amounts to an approximate linearized back extrapolation. If necessary the smoother can be further improved by re-expanding the function f_k in the smoothed point.

The combined filter–smoother algorithm allows the computation of optimal estimates of the track parameters anywhere along the track, using the full information. This has several consequences, which constitute substantial improvements of the simple progressive fit:

(1) It is possible to compute optimal predictions into other detectors from both ends of the track or track segment as well as optimal intersections of the track with a detector, e.g. a RICH.

(2) By removing the measurement \mathbf{m}_k from the smoothed estimate \mathbf{x}_k^n one obtains an optimal estimate \mathbf{x}_k^{n*} of the track in the intersection point with detector k which uses the full information with the exception of \mathbf{m}_k . This estimate can be used for the detection of outliers (see section 5) and for checking and tuning of the detector alignment and resolution. It can be computed successively for all detectors in one single go of the smoother. \mathbf{m}_k can easily be removed from the estimate \mathbf{x}_k^n by an “inverse Kalman filter”. Formally this is a step of the filter with the covariance (or weight) matrix of \mathbf{m}_k taken negative:

$$\mathbf{x}_k^{n*} = \mathbf{x}_k^n + \mathbf{K}_k^{n*} (\mathbf{m}_k - \mathbf{H}_k \mathbf{x}_k^n), \quad (12a)$$

$$\mathbf{K}_k^{n*} = \mathbf{C}_k^n \mathbf{H}_k^T (-\mathbf{V}_k + \mathbf{H}_k \mathbf{C}_k^n \mathbf{H}_k^T)^{-1}, \quad (12b)$$

$$\mathbf{C}_k^{n*} = (\mathbf{I} - \mathbf{K}_k^{n*} \mathbf{H}_k) \mathbf{C}_k^n, \quad (12c)$$

or, in the weighted means formalism:

$$\mathbf{x}_k^{n*} = \mathbf{C}_k^{n*} [(\mathbf{C}_k^n)^{-1} \mathbf{x}_k^n - \mathbf{H}_k^T \mathbf{G}_k \mathbf{m}_k], \quad (12d)$$

$$\mathbf{C}_k^{n*} = [(\mathbf{C}_k^n)^{-1} - \mathbf{H}_k^T \mathbf{G}_k \mathbf{H}_k]^{-1}. \quad (12e)$$

The smoother can also be used for efficient track segment merging: Let us assume that two track segments have been fitted individually in two different detector modules (fig. 1). In order to combine the information from the two segments:

- the optimal estimate (filtered or smoothed) \mathbf{y}_1^m is propagated to the reference surface of \mathbf{x}_n , yielding \mathbf{y}_1' ;
- an updated estimate \mathbf{x}_n' is computed as the weighted mean of \mathbf{x}_n and \mathbf{y}_1' ;
- smoothing proceeds from \mathbf{x}_n' to \mathbf{x}_1'' according to eq. (9) (see fig. 1).

A possible drawback of the filter algorithm is the fact that one needs an initial value of the state vector together with its covariance matrix. This can be obtained by fitting a small number of measurements at the start of the track by a conventional least-squares fit, but this is not an elegant solution. The other possibility is to start with an arbitrary state vector and an “infinite” covariance matrix, i.e. a large multiple of the identity matrix. This is completely in the spirit of the filtering approach, but may lead to numerical instabilities in the computation of the gain matrix, since the infinities have to cancel in order to give a finite gain matrix. This may be difficult on a computer with a short word length.

5. Detection of outliers

In the course of the analysis of an event, track fitting is performed normally after pattern recognition, i.e. after different measurements have been assigned to track candidates. The track fit serves not only to optimally estimate the track parameters but also to assess the quality of this assignment. A commonly used measure of this quality is the global χ^2 of the track. Although the global χ^2 is a powerful test against ghost tracks (random association of coordinates), it loses its

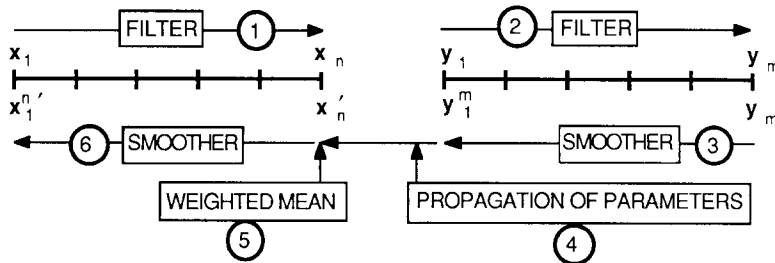


Fig. 1. Track element merging by successive filtering and smoothing.

power against single outliers with increasing number of measurements. Outliers are measurements which do not belong to an otherwise well-defined track. They can be either correlated with the track, for example signals from δ -rays, or uncorrelated, like signals from neighbouring tracks or genuine electronic noise. Track correlated outliers have a distribution around the true track position which depends on the physics of the underlying secondary process and on the properties of the detector. Uncorrelated outliers have, in principle, a flat distribution in the whole detector volume; in practice their distribution is imposed by the selection mechanism of the track finding algorithm.

The residuals of a global fit can be used to find outliers by rejecting measurements with large residuals, but this is only feasible if the residuals ("pulls") are not dominated by multiple scattering. The Kalman filter and smoother offers the possibility to use locally the full information in order to decide whether a measurement is close enough to the track to belong with a high probability to it. A useful decision criterion or measure of "closeness" is the χ^2 of the prediction:

$$\chi_P^2 = \mathbf{r}_k^{k-1T} (\mathbf{R}_k^{k-1})^{-1} \mathbf{r}_k^{k-1}. \quad (13)$$

It can easily be shown that χ_P^2 is equal to the filtered chi-square χ_F^2 :

$$\chi_P^2 = \chi_F^2 = \mathbf{r}_k^T (\mathbf{R}_k)^{-1} \mathbf{r}_k. \quad (14)$$

If \mathbf{m}_k belongs to the track and the covariance matrix of its Gaussian errors is accurately known, χ_F^2 is χ^2 -distributed with m_k degrees of freedom, where m_k is the dimension of \mathbf{m}_k . If \mathbf{m}_k is an outlier, χ_F^2 is noncentrally χ^2 -distributed with m_k degrees of freedom and the noncentral parameter λ_F :

$$\lambda_F = \mathbf{d}^T \mathbf{G}_k \mathbf{R}_k \mathbf{G}_k \mathbf{d} = \mathbf{d}^T \mathbf{G}_k (\mathbf{I} - \mathbf{H}_k \mathbf{K}_k) \mathbf{d}, \quad (15)$$

where \mathbf{d} is the offset of the outlier with respect to the real track position:

$$\mathbf{m}_k = \mathbf{H}_k \mathbf{x}_{k,i} + \mathbf{d} + \boldsymbol{\epsilon}_k. \quad (16)$$

If χ_F^2 is larger than a given bound c , the measurement is rejected as an outlier. If c is chosen as the $(1 - \alpha)$ quantile of the appropriate χ^2 distribution, the probability to reject a good measurement is equal to α . The probability of rejecting an outlier (the power of the χ^2 test), as a function of m_k , α and λ_F , is computed in the appendix. For a method to estimate accurately the covariance matrix of $\boldsymbol{\epsilon}_k$ see, e.g., ref. [10]. A similar χ^2 test can be performed by using the smoothed residual:

$$\chi_S^2 = \mathbf{r}_k^{nT} (\mathbf{R}_k^n)^{-1} \mathbf{r}_k^n. \quad (17)$$

If \mathbf{x}_k^{n*} is the smoothed estimate of $\mathbf{x}_{k,i}$ without using \mathbf{m}_k (see eq. (12)) we have clearly (see also eq. (8a)):

$$\mathbf{x}_k^n = \mathbf{x}_k^{n*} + \mathbf{K}_k^n (\mathbf{m}_k - \mathbf{H}_k \mathbf{x}_k^{n*}), \quad (18a)$$

$$\mathbf{K}_k^n = \mathbf{C}_k^n \mathbf{H}_k^T \mathbf{G}_k. \quad (18b)$$

If \mathbf{m}_k is the only outlier, \mathbf{x}_k^{n*} is unbiased and

$$E\{\mathbf{r}_k^n\} = \mathbf{d} - \mathbf{H}_k \mathbf{C}_k^n \mathbf{H}_k^T \mathbf{G}_k \mathbf{d} = \mathbf{R}_k^n \mathbf{G}_k \mathbf{d}. \quad (19)$$

(The expectation is taken over all outliers with the same \mathbf{d} .) χ_S^2 is again noncentrally χ^2 -distributed with the noncentral parameter λ_S :

$$\lambda_S = \mathbf{d}^T \mathbf{G}_k \mathbf{R}_k^n \mathbf{G}_k \mathbf{d}. \quad (20)$$

If we compare λ_S and λ_F , we find that

$$\begin{aligned} \lambda_S - \lambda_F &= \mathbf{d}^T \mathbf{G}_k (\mathbf{R}_k^n - \mathbf{R}_k) \mathbf{G}_k \mathbf{d} \\ &= \mathbf{d}^T \mathbf{G}_k \mathbf{H}_k \mathbf{A}_k (\mathbf{C}_{k+1}^k - \mathbf{C}_{k+1}^n) \mathbf{A}_k^T \mathbf{H}_k^T \mathbf{G}_k \mathbf{d} \geq 0, \end{aligned} \quad (21)$$

since it is easily shown by induction that $\mathbf{C}_{k+1}^k - \mathbf{C}_{k+1}^n$ is positive definite. Therefore the test on χ_S^2 is always more powerful than the test on $\chi_F^2 = \chi_P^2$. This is in agreement with the intuitive reasoning. Also, the global χ^2 being the sum of all filtered chi-squares, the test on the global χ^2 is always less powerful than the test on χ_F^2 . This means that a search for possible outliers should be performed during smoothing, when the full information on the track parameters is available. It is clear that outliers with large \mathbf{d} are found more frequently than the ones with small \mathbf{d} ; fortunately the latter are less harmful, as they introduce less bias into the final estimate.

If \mathbf{m}_k is an outlier and has to be removed permanently from the track, one may just continue smoothing with \mathbf{x}_k^{n*} and \mathbf{C}_k^{n*} instead of with \mathbf{x}_k^n and \mathbf{C}_k^n . However, this does not update the estimates \mathbf{x}_j^n with $j > k$. If the whole track has to be updated, the filter has to be recomputed, starting from \mathbf{x}_k^{k-1} and without using \mathbf{m}_k , followed by smoothing back over the whole track.

6. Application of the Kalman filter to a vertex fit

We show now that the fast vertex fit proposed in ref. [2] is a special case of a nonlinear Kalman filter. Initially the state vector consists only of the prior information about the vertex position, \mathbf{x}_0 and $\mathbf{C}_0 = \text{cov}\{\mathbf{x}_0\}$. For each 5-vector \mathbf{p}_k of fitted track parameters the state vector is augmented by the 3-vector \mathbf{q}_k of momentum of track k at the vertex. The system equation is particularly simple:

$$\mathbf{x}_k = \mathbf{x}_{k-1}. \quad (22)$$

The measurement equation is – in the presence of a magnetic field – nonlinear:

$$\mathbf{p}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{q}_k) + \boldsymbol{\epsilon}_k. \quad (23)$$

As usual, we linearize \mathbf{h}_k in some point $(\mathbf{x}_{k,0}, \mathbf{q}_{k,0})$:

$$\begin{aligned} \mathbf{h}_k(\mathbf{x}_k, \mathbf{q}_k) &= \mathbf{h}_k(\mathbf{x}_{k,0}, \mathbf{q}_{k,0}) + \mathbf{A}_k(\mathbf{x}_k - \mathbf{x}_{k,0}) \\ &\quad + \mathbf{B}_k(\mathbf{q}_k - \mathbf{q}_{k,0}) \\ &= \mathbf{c}_{k,0} + \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{q}_k. \end{aligned} \quad (24)$$

Since there is usually no prior information about \mathbf{q}_k , we assign an “infinite” covariance matrix to the “predicted” vector \mathbf{q}_k^{k-1} :

$$\mathbf{D}_k^{k-1} = \text{cov}\{\mathbf{q}_k^{k-1}\} = (1/\delta)\mathbf{I}, \quad \delta \text{ small.} \quad (25)$$

Then the prediction equations look as follows:

$$\mathbf{x}_k^{k-1} = \mathbf{x}_{k-1}, \quad \mathbf{q}_k^{k-1} = \mathbf{q}_{k,0}, \quad (26)$$

$$\mathbf{C}_k^{k-1} = \mathbf{C}_{k-1}, \quad \mathbf{D}_k^{k-1} = (1/\delta)\mathbf{I}.$$

We derive now the filter equations in the weighted means formulation:

$$\begin{bmatrix} \mathbf{x}_k \\ \mathbf{q}_k \end{bmatrix} = \begin{bmatrix} \mathbf{C}_k & \mathbf{E}_k \\ \mathbf{E}_k^T & \mathbf{D}_k \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{C}_{k-1}^{-1}\mathbf{x}_{k-1} + \mathbf{A}_k^T \mathbf{G}_k (\mathbf{p}_k - \mathbf{c}_{k,0}) \\ (\mathbf{D}_k^{k-1})^{-1} \mathbf{q}_k^{k-1} + \mathbf{B}_k^T \mathbf{G}_k (\mathbf{p}_k - \mathbf{c}_{k,0}) \end{bmatrix}, \quad (27a)$$

$$\begin{bmatrix} \mathbf{C}_k & \mathbf{E}_k \\ \mathbf{E}_k^T & \mathbf{D}_k \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{k-1}^{-1} + \mathbf{A}_k^T \mathbf{G}_k \mathbf{A}_k & \mathbf{A}_k^T \mathbf{G}_k \mathbf{B}_k \\ \mathbf{B}_k^T \mathbf{G}_k \mathbf{A}_k & \mathbf{B}_k^T \mathbf{G}_k \mathbf{B}_k \end{bmatrix}^{-1}. \quad (27b)$$

After some matrix algebra and taking the limit $\delta \rightarrow 0$ we obtain the following results which are identical to the ones in ref. [2]:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{C}_k \left[\mathbf{C}_{k-1}^{-1} \mathbf{x}_{k-1} + \mathbf{A}_k^T \mathbf{G}_k^B (\mathbf{p}_k - \mathbf{c}_{k,0}) \right], \\ \mathbf{q}_k &= \mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k (\mathbf{p}_k - \mathbf{c}_{k,0} - \mathbf{A}_k \mathbf{x}_k), \\ \mathbf{C}_k &= (\mathbf{C}_{k-1}^{-1} + \mathbf{A}_k^T \mathbf{G}_k^B \mathbf{A}_k)^{-1}, \end{aligned} \quad (28)$$

$$\mathbf{D}_k = \mathbf{W}_k + \mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k \mathbf{A}_k \mathbf{C}_k \mathbf{A}_k^T \mathbf{G}_k \mathbf{B}_k \mathbf{W}_k,$$

$$\mathbf{E}_k = -\mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k \mathbf{A}_k \mathbf{C}_k,$$

with

$$\mathbf{W}_k = (\mathbf{B}_k^T \mathbf{G}_k \mathbf{B}_k)^{-1},$$

$$\mathbf{G}_k^B = \mathbf{G}_k - \mathbf{G}_k \mathbf{B}_k \mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k,$$

$$\text{cov}\{\mathbf{x}_k\} = \mathbf{C}_k, \quad \text{cov}\{\mathbf{q}_k\} = \mathbf{D}_k, \quad \text{cov}\{\mathbf{x}_k, \mathbf{q}_k\} = \mathbf{E}_k.$$

The χ^2 increment is given by (see eq. (8b)):

$$\begin{aligned} \chi_+^2 &= (\mathbf{p}_k - \mathbf{c}_{k,0} - \mathbf{A}_k \mathbf{x}_k - \mathbf{B}_k \mathbf{q}_k)^T \mathbf{G}_k \\ &\quad \times (\mathbf{p}_k - \mathbf{c}_{k,0} - \mathbf{A}_k \mathbf{x}_k - \mathbf{B}_k \mathbf{q}_k) \\ &\quad + (\mathbf{x}_k - \mathbf{x}_{k-1})^T \mathbf{C}_{k-1}^{-1} (\mathbf{x}_k - \mathbf{x}_{k-1}), \\ \chi_k^2 &= \chi_{k-1}^2 + \chi_+^2. \end{aligned} \quad (29)$$

If necessary, the linear expansion can now be repeated in the new point:

$$\mathbf{x}_{k,0} = \mathbf{x}_k, \quad \mathbf{q}_{k,0} = \mathbf{q}_k,$$

and the filter can be recomputed, until there is no significant change either in the χ^2 or in the estimate.

Since there is no process noise, the smoother is extremely simple:

$$\begin{aligned} \mathbf{x}_k^n &= \mathbf{x}_n, \\ \mathbf{q}_k^n &= \mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k (\mathbf{p}_k - \mathbf{c}_{k,0} - \mathbf{A}_k \mathbf{x}_n), \\ \mathbf{C}_k^n &= \mathbf{C}_n, \\ \mathbf{D}_k^n &= \mathbf{W}_k + \mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k \mathbf{A}_k \mathbf{C}_n^T \mathbf{A}_k^T \mathbf{G}_k \mathbf{B}_k \mathbf{W}_k, \\ \mathbf{E}_k^n &= -\mathbf{W}_k \mathbf{B}_k^T \mathbf{G}_k \mathbf{A}_k \mathbf{C}_n^n. \end{aligned} \quad (30)$$

If there is a significant change in the smoothed vertex position, it may be worthwhile to recompute the derivative matrices \mathbf{A}_k and \mathbf{B}_k .

7. Detection of secondary vertices

We assume that only a few tracks originate possibly from a secondary vertex, so that the estimated position of the primary vertex has no noticeable bias. Again, the filtered or smoothed residuals can be used to decide whether or not a particular track really does belong to the primary vertex. The residuals and their covariance matrices have the following form (see eqs. (8) and (9)):

$$\begin{aligned} \mathbf{r}_k &= \mathbf{p}_k - \mathbf{c}_{k,0} - \mathbf{A}_k \mathbf{x}_k - \mathbf{B}_k \mathbf{q}_k, \\ \mathbf{R}_k &= \mathbf{V}_k (\mathbf{G}_k^B - \mathbf{G}_k^B \mathbf{A}_k \mathbf{C}_k \mathbf{A}_k^T \mathbf{G}_k^B) \mathbf{V}_k, \end{aligned} \quad (31)$$

$$\begin{aligned} \mathbf{r}_k^n &= \mathbf{p}_k - \mathbf{c}_{k,0} - \mathbf{A}_k \mathbf{x}_n - \mathbf{B}_k \mathbf{q}_k^n, \\ \mathbf{R}_k^n &= \mathbf{V}_k (\mathbf{G}_k^B - \mathbf{G}_k^B \mathbf{A}_k \mathbf{C}_n \mathbf{A}_k^T \mathbf{G}_k^B) \mathbf{V}_k. \end{aligned}$$

Since \mathbf{R}_k and \mathbf{R}_k^n are singular the filtered chi-square χ_F^2 and the smoothed chi-square χ_S^2 have to be computed in the following way (see eq. (29)):

$$\chi_F^2 = \mathbf{r}_k^T \mathbf{G}_k \mathbf{r}_k + (\mathbf{x}_k - \mathbf{x}_{k-1})^T \mathbf{C}_{k-1}^{-1} (\mathbf{x}_k - \mathbf{x}_{k-1}), \quad (32a)$$

$$\chi_S^2 = \mathbf{r}_k^{nT} \mathbf{G}_k \mathbf{r}_k^n + (\mathbf{x}_n - \mathbf{x}_k^{n*})^T (\mathbf{C}_k^{n*})^{-1} (\mathbf{x}_n - \mathbf{x}_k^{n*}), \quad (32b)$$

where \mathbf{x}_k^{n*} is the smoothed estimate \mathbf{x}_n with the track \mathbf{p}_k removed. It is obtained by the inverse Kalman filter (see eq. (12)):

$$\mathbf{C}_k^{n*} = (\mathbf{C}_n^{-1} - \mathbf{A}_k^T \mathbf{G}_k^B \mathbf{A}_k)^{-1}, \quad (33)$$

$$\mathbf{x}_k^{n*} = \mathbf{C}_k^{n*} [\mathbf{C}_n^{-1} \mathbf{x}_n - \mathbf{A}_k^T \mathbf{G}_k^B (\mathbf{p}_k - \mathbf{c}_{k,0})].$$

If \mathbf{p}_k belongs to the primary vertex, χ_F^2 and χ_S^2 are χ^2 -distributed with 2 degrees of freedom. If \mathbf{p}_k originates from a secondary vertex \mathbf{z} not too far from the primary vertex \mathbf{x} , we can write in linear approximation:

$$\mathbf{p}_k = \mathbf{h}_k(\mathbf{x}, \mathbf{q}_{k,t}) + \mathbf{A}_k(\mathbf{z} - \mathbf{x}) + \epsilon_k.$$

We may choose \mathbf{z} in such a way that $\mathbf{d} = \mathbf{z} - \mathbf{x}$ is orthogonal to $\mathbf{q}_{k,t}$. The impact parameter of track k is then given by $|\mathbf{d}|$ and the offset of \mathbf{p}_k with respect to the primary vertex by

$$\mathbf{E}\{\mathbf{p}_k - \mathbf{c}_{k,0} - \mathbf{A}_k \mathbf{x} - \mathbf{B}_k \mathbf{q}_{k,t}\} = \mathbf{A}_k \mathbf{d}.$$

In analogy to eqs. (15) and (20) we can now compute the noncentral parameters of χ_F^2 and χ_S^2 :

$$\lambda_F = d^T A_k^T (G_k^B - G_k^B A_k C_k A_k^T G_k^B) A_k d, \quad (34a)$$

$$\lambda_S = d^T A_k^T (G_k^B - G_k^B A_k C_n A_k^T G_k^B) A_k d. \quad (34b)$$

With λ_F and λ_S we can in turn compute the power of the respective χ^2 test (see appendix).

Acknowledgements

I thank M. Regler for his lively interest in this work, for many useful discussions and improvements of the manuscript. I thank P. Laurikainen for drawing my attention to his early ideas about the application of filtering techniques in the analysis of bubble chamber experiments. I also thank P. Billoir for rousing my interest in recursive methods of track fitting and for many useful comments.

Appendix

The power of a χ^2 test

Let us assume that χ^2 is distributed according to a χ^2 distribution $g(n; x)$ if the hypothesis H_0 is true and according to a noncentral χ^2 distribution $g(n, \lambda; x)$ if H_1 is true. Let c be equal to the $(1 - \alpha)$ quantile of $g(n; x)$.

Then the power function $p(n, \lambda, \alpha)$ of the χ^2 test with respect to the hypothesis H_1 is given by:

$$p(n, \lambda, \alpha) = \int_c^\infty g(n, \lambda; x) dx.$$

In order to compute p , we expand $g(n, \lambda; x)$ into a (uniformly convergent) series of central χ^2 probability density functions [11]:

$$g(n, \lambda; x) = \sum_{k=0}^{\infty} \frac{\exp(-\lambda/2) \lambda^k}{2^k k!} g(n + 2k; x).$$

Then we have:

$$p(n, \lambda, \alpha) = \sum_{k=0}^{\infty} \frac{\exp(-\lambda/2) \lambda^k}{2^k k!} \int_c^\infty g(n + 2k; x) dx.$$

By partial integration we obtain the following recursion:

$$\int_c^\infty g(2j; x) dx = \int_c^\infty g(2j - 2; x) dx + \frac{\exp(-c/2) c^{j-1}}{2^{j-1} (j-1)!}.$$

Hence we have:

$$\int_c^\infty g(2n + k; x) dx = \alpha + \sum_{i=1}^k \frac{\exp(-c/2) c^i}{2^i i!},$$

and:

$$p(n, \lambda, \alpha) = \alpha + \exp(-\lambda/2) \exp(-c/2)$$

$$\times \sum_{k=1}^{\infty} \frac{\lambda^k}{2^k k!} \sum_{i=1}^k \frac{c^i}{2^i i!}.$$

References

- [1] P. Billoir, Nucl. Instr. and Meth. 225 (1984) 352.
- [2] P. Billoir, R. Frühwirth and M. Regler, Nucl. Instr. and Meth. A241 (1985) 115.
- [3] DELPHI Technical Proposal, CERN/LEPC/83-3.
- [4] P. Laurikainen, Report Series in Physics 35 (University of Helsinki, 1971).
- [5] A. Gelb (ed.), Applied Optimal Estimation (MIT Press, Cambridge and London, 1975).
- [6] K. Brammer und G. Siffling, Kalman-Bucy-Filter (R. Oldenbourg, München, Wien, 1975).
- [7] S.M. Bozic, Digital and Kalman Filtering (Edward Arnold, London, 1979).
- [8] A. Jazwinski, Stochastic Processes and Filtering Theory (Academic Press, New York, San Francisco, London, 1970).
- [9] M. Regler, Acta Phys. Austr. 49 (1978) 37; Formulae and Methods in Experimental Data Evaluation, vol. 2 (European Physical Society, 1984).
- [10] R. Frühwirth, Nucl. Instr. and Meth. A243 (1985) 173.
- [11] S. Kotz, N.L. Johnson and D.W. Boyd, Ann. Math. Statist. 38 (1967) 838.